

PATENT COOPERATION TREATY

WO 99/62002
PCT/FI98/00441

PCT

From the INTERNATIONAL BUREAU

NOTICE INFORMING THE APPLICANT OF THE COMMUNICATION OF THE INTERNATIONAL APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

To:

BERGGREN OY AB
P.O. Box 16
FIN-00101 Helsinki
FINLANDE

Berggren Oy Ab

14 -12- 1999

Date of mailing (day/month/year) 02 December 1999 (02.12.99)		IMPORTANT NOTICE	
Applicant's or agent's file reference 46682 /JK			
International application No. PCT/FI98/00441	International filing date (day/month/year) 27 May 1998 (27.05.98)	Priority date (day/month/year)	
Applicant BECKS, Ari et al			

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:
EP,JP,US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:
EA,NO

The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on 02 December 1999 (02.12.99) under No. WO 99/62002

REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, a demand for international preliminary examination must be filed with the competent International Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))

If the applicant wishes to proceed with the international application in the national phase, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer J. Zahra
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 46682	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/ISA/416)	
International application No. PCT/FI98/00441	International filing date (day month year) 27.05.1998	Priority date (day month year) -
International Patent Classification (IPC) or national classification and IPC G06F 17/28		
Applicant BECKS Ari et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 4 sheets, including this cover sheet.

☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of _____ sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand 27.12.1999	Date of completion of this report 11.09.2000
Name and mailing address of the IPEA/SE Patent- och registreringsverket Box 5055 S-102 42 STOCKHOLM Facsimile No. 08-667 72 88	Authorized officer Oskar Pihlgren/LR Telephone No. 08-782 25 00

Form PCT/ISA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/FI98/00441

I. Basis of the report

1. This report has been drawn on the basis of *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.)*:

- ☒ the international application as originally filed.
- ☐ the description, pages _____, as originally filed,
 pages _____, filed with the demand,
 pages _____, filed with the letter of _____,
 pages _____, filed with the letter of _____.
- ☐ the claims, Nos. _____, as originally filed,
 Nos. _____, as amended under Article 19,
 Nos. _____, filed with the demand,
 Nos. _____, filed with the letter of _____,
 Nos. _____, filed with the letter of _____.
- ☐ the drawings, sheets/fig _____, as originally filed,
 sheets/fig _____, filed with the demand,
 sheets/fig _____, filed with the letter of _____,
 sheets/fig _____, filed with the letter of _____.

2. The amendments have resulted in the cancellation of:

- ☐ the description, pages _____
- ☐ the claims, Nos. _____
- ☐ the drawings, sheets/fig _____

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the supplemental Box (Rule 70.2(c)).

4. Additional observations, if necessary:

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/FI98/00441

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Claims	<u>1-15</u>	YES
	Claims		NO
Inventive step (IS)	Claims	<u>1-15</u>	YES
	Claims		NO
Industrial applicability (IA)	Claims	<u>1-15</u>	YES
	Claims		NO

2. Citations and explanations

The claimed invention

The claimed invention relates to a system and method for machine translation of a character string in a first language into a second language. The character string to be translated is divided into model segments on the basis of a first rule. An identified structural segment is compared with model segments stored in a memory whereby one is selected. An equivalent segment is chosen that is logically connected to the model segment to represents a proper translation.

The present invention is characterized in that the identification of model segments in the character string to be translated, is based on identification of an intermediate word and/or suffix.

Cited documents

The following documents were cited in the International Search Report:

D1: EP 0805403 A2

D2: EP 0262938 A1

Statement

D2 relates to a language translation system for translating phrases from a first language into a second language. A memory stores a collection of phrases in the second language (see fig 1). Phrases input in the first language are each characterized on the basis of one or more keywords, and the corresponding phrase in the second language is output.

.../...

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/FI98/00441

Supplemental Box

(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: V.

D1, which is considered to be the prior-art most closely related to the present invention, relates to a translating apparatus and method wherein a first language sentence is divided (4) into syntax units consisting of predetermined units of sentence structure such as clauses and phrases. Stored examples (3, 5) most similar to these syntax units are detected (6) using probability (11) models taking into account grammatical attributes of the syntax units and of the examples. The syntax units are translated (7) on the basis of the detected examples, and the results of the translation of the syntax units are compounded to generate a second language sentence.

From D1 it is prior known to divide a character string (sentence) of a first language into segments, and to translate this character string on the basis of stored examples into a second language. The difference between the invention and D1 is that the identification of a structural segment according to the invention is based on the identification of an intermediate word and/or suffix. This is not mentioned in D1 and it is not considered to be obvious to a person skilled in the art.

With reference to the reasons stated above, the invention according to claims 1-15 is novel and considered to involve an inventive step.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



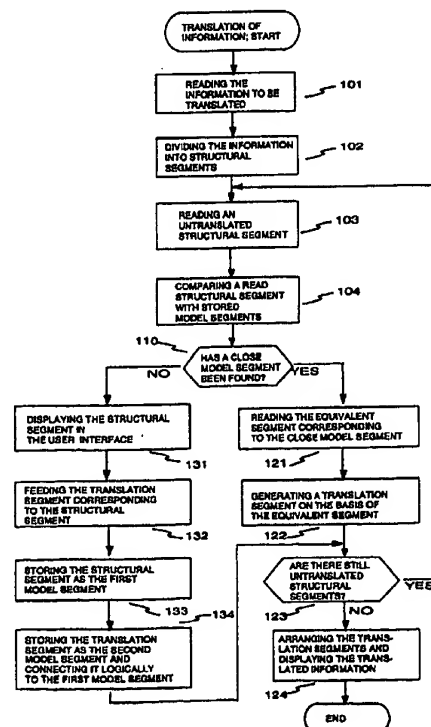
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 17/28	A1	(11) International Publication Number: WO 99/62002 (43) International Publication Date: 2 December 1999 (02.12.99)
(21) International Application Number: PCT/FI98/00441 (22) International Filing Date: 27 May 1998 (27.05.98) (71)(72) Applicants and Inventors: BECKS, Ari [FI/FI]; Osu- uskunnantie 100 C 3, FIN-00660 Helsinki (FI). HEIKKILÄ, Simo, Sakari [FI/FI]; Upseerinkatu 1-5 A 2, FIN-15700 Lahti (FI). (74) Agent: BERGGREN OY AB; P.O. Box 16, FIN-00101 Helsinki (FI).		(81) Designated States: JP, NO, US, Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published With international search report. In English translation (filed in Finnish).

(54) Title: A METHOD AND ARRANGEMENT FOR TRANSLATION OF INFORMATION

(57) Abstract

The invention relates to a method and an arrangement for translating information given as a character string in a first language into a character string in a second language. The invention is advantageously implemented in machine translation of text information. One idea of the invention is to divide the information to be translated into structural segments (102) and to perform the translation by structural segments (122). The translation is done on the basis of model segments and rules stored in the knowledge base. The data included in the knowledge base are advantageously increased in such a way that, whenever necessary in the translating process, the user feeds translations (132) of new model segments over the user interface, these translations being subsequently stored as model segments in the knowledge base (133, 134). Owing to the solution of the invention, the translating equipment requires less memory capacity and a lower processor speed. In addition, it requires substantially less programming and the operation of the equipment can be developed without software updating.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

A method and arrangement for translation of information

5 The invention relates to a method and an arrangement for translating information given as a character string in a first language into a character string in a second language. The invention is advantageously implemented in machine translation of text information.

10 There are previously known methods for linguistically based machine translation of text information. In these methods, the syntax of each language is exactly programmed, so that each language will require a program algorithm of its own. For the storage of vocabularies in different languages, a centralised high-capacity translation memory is used. The EuroTra translation system of the European Union can be mentioned as an example of such a method. Such previously known methods
15 have a number of drawbacks. Exact syntax programming requires most extensive programming operations. Such a syntax algorithm, as well as the necessary translation memory, require a large memory space in the database. Since a translation method operating in this manner is complex, translating within a reasonable time requires an extremely high-powered computer. Due to these
20 shortcomings, the equipment suitable for translation is expensive. Known methods also involve the drawback that updating of the translation algorithm requires programming and updating of the computer program each time.

25 The object of the present invention is to provide a solution for the translation of information which enables the prior art inconveniences described above to be overcome.

30 One idea of the invention is to divide the information to be translated into structural segments and to do the translation by structural segments. The translation is performed on the basis of model segments and rules stored in the knowledge base. The data contained in the knowledge base are advantageously increased so that, in the process of translating, whenever necessary, the user is asked to provide translations of new structural segments over a user interface, these translations being
35 subsequently stored as model segments in the knowledge base. Owing to the solution provided by the invention, the translating equipment requires a smaller memory capacity and a lower processor speed. Moreover, far less programming is required and the operation of the equipment can be developed without program updating.

The method of the invention for machine translation of information given as a character string in a first language into a character string in a second language is characterised by

- 5 - storing model segments in the form of character strings in the first language in the knowledge base and, logically connected to these, model segments in the form of character strings in the second language,
- identifying a structural segment in the character string of said first language following a first rule,
- 10 - comparing said identified structural segment with model segments in the form of character strings in the first language stored according to a second rule,
- striving to select one model segment on the basis of said comparison,
- reading a model segment in the form of a character string in the second language logically connected to the selected model, i.e. equivalent segment and
- 15 - translating said structural segment into said translation segment in the form of a character string in the second language on the basis of said equivalent segment and a third rule.

The arrangement of the invention for translating information given as a character string in a first language into a character string in a second language is characterised in comprising

- 20 - knowledge base means for storing model segments in the form of said character strings in the first language and, in logical connection with these, for storing equivalent segments in the form of character strings in the second language, and for
- 25 storing a first, second and third rule,
- means for identifying structural segments in said information given as a character string in the first language following a first rule,
- means for comparing said identified structural segment with the stored model segments in the form of character strings in the first language following a second
- 30 rule,
- means for selecting one model segment on the basis of said comparison,
- means for reading a model segment, i.e. equivalent segment, in the form of a character string in the second language, logically connected to the selected model, in said knowledge base means and
- 35 - means for translating said structural segment into said translation segment in the form of a character string in the second language on the basis of said equivalent segment and the third rule, said translation segment representing the information to be given in said second language.

then stored. The stored information can be further displayed, e.g. on a screen, or printed out e.g. on paper or a disc, block 124.

If no model segment close to the structural segment is found in the knowledge base
5 in block 110, this particular structural segment is displayed over a user interface
means, i.e. a display screen, block 131. The user then feeds the translation of the
structural segment, i.e. the equivalent segment, block 132. The structural segment
and the equivalent segment are stored for future use as model segments in the
knowledge base, blocks 133, 134. After this the process proceeds to block 123 to
10 continue as explained above. In this case, the equivalent segment is usually directly
a translation segment, if the user has been asked to give the translation of the
structural segment in the form of the original information. Thus the operation of
block 122 is not indispensable in this case.

15 Said first rule, by which the structural segments are identified, can be based for
instance on the identification of "intermediate words" or cases. Intermediate words
are for instance prepositions and particles, which usually form standard character
strings. Thus, they can be identified by simply comparing the character strings
forming each word e.g. with the above known character strings forming an
20 intermediate word. The identification of cases can be performed e.g. with the aid of
suffixes by comparing the last characters of the words with known suffixes. As well
known, the character strings forming a word can be separated by means of
punctuation. Since a structural segment may advantageously comprise several
words, it may also include one or more punctuation marks.

25 In its most straightforward version, said second rule, by which a structural segment
is compared with the model segments, may imply similarity. In this case, exactly the
same model segment as the present structural segment to be translated is searched in
the knowledge base. However, considering the memory space required for the
30 knowledge base, it is preferable not to store the different cases of e.g. the model
segment separately in the knowledge base, but to identify also a model segment
having a different case following the second rule. In this situation, the equivalent
segment logically connected to the model segment should also be put in the case
needed in order to generate a translation segment. This is done according to the third
35 rule, which consequently covers information about the cases of the language in
question.

In many cases, said fourth rule, by which the translation segments are arranged in translated sentences, implies placing the translation segments into the same order in which the structural segments to be translated were in the first language. Yet this order may depend on the language, and hence also said fourth rule is language-specific.

In the storage of the model segments, a type identifier of the model segment can also be advantageously stored. In this case, the type identifier is stored in logical connection with each model segment. If type identifiers are used, various rules can be applied the identification and translation of the structural segment on the basis of the model segment, depending on the type of the structural segment. Types of structural segments are e.g. the object of an action, a proper name, a verb, a place word, an adjective or an idiom. If type identifiers are used, the user is also asked to indicate the type to which the particular structural segment and its translation pertain as the structural segment is translated.

One idea of the invention is to update the knowledge base in the interactively operated translation process. It should be noted that the updating of the knowledge base is not necessarily confined to the storage of new model or equivalent segments, but the rules mentioned above can also be advantageously updated. The updating is then performed e.g. in connection with the translation of a new structural segment fed by the user by identifying the regularity of the input translation.

The translation of one piece of information from a first language into a second language has been described above. The previous updatings of the knowledge base are advantageously utilised in the translation of the subsequent pieces of information. Thus, the process of the invention for translating successive first and second pieces of information may comprise e.g. the following steps:

- reading first information given as a character string in the first language,
- performing the translation of the first information given as a character string in said first language on the basis of data in the knowledge base into first information given as a character string in the second language to the extent this is feasible in terms of the data available in the knowledge base,
- determining the additional data required to complete the translation of the first information given as a character string in the first language into first information given as a character string in the second language,
- feeding said additional data in the knowledge base with a view to update the knowledge base,

In many cases, said fourth rule, by which the translation segments are arranged in translated sentences, implies placing the translation segments into the same order in which the structural segments to be translated were in the first language. Yet this order may depend on the language, and hence also said fourth rule is language-specific.

In the storage of the model segments, a type identifier of the model segment can also be advantageously stored. In this case, the type identifier is stored in logical connection with each model segment. If type identifiers are used, various rules can be applied the identification and translation of the structural segment on the basis of the model segment, depending on the type of the structural segment. Types of structural segments are e.g. the object of an action, a proper name, a verb, a place word, an adjective or an idiom. If type identifiers are used, the user is also asked to indicate the type to which the particular structural segment and its translation pertain as the structural segment is translated.

One idea of the invention is to update the knowledge base in the interactively operated translation process. It should be noted that the updating of the knowledge base is not necessarily confined to the storage of new model or equivalent segments, but the rules mentioned above can also be advantageously updated. The updating is then performed e.g. in connection with the translation of a new structural segment fed by the user by identifying the regularity of the input translation.

The translation of one piece of information from a first language into a second language has been described above. The previous updatings of the knowledge base are advantageously utilised in the translation of the subsequent pieces of information. Thus, the process of the invention for translating successive first and second pieces of information may comprise e.g. the following steps:

- reading first information given as a character string in the first language,
- performing the translation of the first information given as a character string in said first language on the basis of data in the knowledge base into first information given as a character string in the second language to the extent this is feasible in terms of the data available in the knowledge base,
- determining the additional data required to complete the translation of the first information given as a character string in the first language into first information given as a character string in the second language,
- feeding said additional data in the knowledge base with a view to update the knowledge base,

- finishing the translation of the first information given as a character string in the first language into first information given as a character string in the second language,
- storing said first information given as a character string in the second language,
- 5 - storing the second information given as a character string in the first language,
- performing the translation of the second information given as a character string in said first language on the basis of said updated data in the knowledge base into second information given as a character string in the second language.

10 Figure 2 is a block diagram of a device arrangement of the invention for the translation of information. The arrangement comprises a disc station 21, a display screen 22 and a keyboard 23 as interface means connected to processor 20. By means of the disc station, information to be translated can be fed from the disc to the device and the translated information can be stored on the disc for use in other
15 devices. The information in question can be transferred between the device and other data processing equipment also over a bus I/O. Display screen 22 can be used to display such structural segments to the user for which no translation is found in the knowledge base. The user can feed the translation of such a structural segment by using keyboard 23. The interface means mentioned above can also be used in the
20 revision and correction of translated information.

The device shown in figure 2 also comprises an electric memory 24 for temporary storage of structural segments and translation segments, among other things. In addition, the device comprises a mass storage 25 for the storage of the knowledge
25 base, i.e. model segments, type identifiers and rules, as well as programs. For instance a hard disc drive or an optical disc drive can be used as a mass storage. The components mentioned above can be provided by making previously known computer components operate in accordance with the invention using special software. Character strings and other data are advantageously transferred as electric
30 signals between the components.

The implementation of the invention is by no means confined to the components described above, by contrast, the arrangement of the invention can have many different configurations, which this description enables a person skilled in the art to
35 design.

Figure 3 illustrates an English sentence divided into structural segments 31, 32, 33 and 34. As shown in the figure, a structural segment typically comprises successive

closely related words in a sentence. Thus a structural segment often includes a punctuation mark separating the words as well.

Figure 4 illustrates the translation of the first structural segment of the sentence appearing in figure 3 with the aid of one solution of the invention. In the figure, the structural segment 42 to be translated is stored in translation memory 41 and this structural segment is compared with the model segments stored in knowledge base 44. In the case illustrated in figure 4, this particular structural segment has been previously stored in the knowledge base as model segment 45, which is found in the comparison. If, for instance, the present information is to be translated into Finnish, the Finnish model segment 46 logically connected to the English model segment mentioned above is read in the knowledge base. In figure 4 the double line connecting model segments 45 and 46 illustrates a logical connection. When the Finnish model segment has been read it is stored as a translation segment in the translation memory.

Figure 5 illustrates the translation of the second structural segment shown in figure 3 with the aid of a solution of the invention. In this case, no English structural segment to be translated nor any Finnish equivalent segment has been previously stored as a model segment in the knowledge base. In this case, structural segment 52 to be translated, stored in translation memory 51, is compared with the model segments in the knowledge base, and if the desired equivalent segment is not found in the knowledge base, the structural segment 58 to be translated is shown on the display screen of interface 57. After this, the user feeds the translation 59 of structural segment 58 over the interface in knowledge base 54. In this manner, an English and a Finnish model segment are stored in logical connection in the knowledge base. Then the Finnish translation of the structural segment is stored as a translation segment 53 in translation memory 51.

Should the structural segments mentioned above reappear in the input information, corresponding model and equivalent segments will be found in the knowledge base, and there will be no need to ask the user to repeat them. If, however, the following input information contains the sentence "we have expanded our operation largely in Finland", "largely" would be a new structural segment. If no close model segment has been previously stored in the knowledge base, the user is asked to give the translation of it and "largely" is stored as a model segment in the knowledge base, and in logical connection with this, also the input translation fed by the user.

It should be noted that the operation of the equipment can be arranged so that the translation process is first performed by machine for the entire information to the extent allowed by the model segments stored in the knowledge base. After this the user can feed the necessary translations of new structural segments in the knowledge base. Such an arrangement has the advantage of the user not having to stay by the computer waiting for the translation process to be completed, but he/she may update the knowledge base with one single input at any suitable moment.

The model segments can be stored in the knowledge base as pairs of segments, specific pairs of model segments being stored for each language pair. Another way of proceeding is to logically connect model segments in several languages, so that the same model segments can be used as such in the translation of several language pairs. In this case, the model segments of each language can be fed as an input in the knowledge base each time they appear for the first time in the language in question. When input information is then fed in the knowledge base during the translation of one language pair, the information contained in the knowledge base will automatically increase also in the other language pairs.

The solution of the invention is not language-specific on principle, but can be applied to any language pair. Nor is the implementation of the invention restricted to "natural" languages used in ordinary communication, since it can be used to translate any language consisting of character strings into a second language consisting of character strings. Programming languages and data exchange protocols may be mentioned as examples of such other languages.

The solution of the invention has many advantages over prior art. Its operation requires but little language-specific knowledge for the division of the language into structural segments. A second advantage of the solution consists in additional information being collected in the memory during the process, so that the device "learns" new pairs of model segments and rules. Thus, with a straightforward configuration and a small amount of programming and updating it is possible to provide an efficient means for machine translation.

The solution of the invention is well adapted for use in situations where the arrangement of the invention is used to meet the needs of several users. In this case, the arrangement preferably comprises several interfaces, which may communicate with the knowledge base e.g. over a data transmission network. The knowledge base can then preferably be decentralised in such a way that the first, i.e. the main

knowledge base can be used by a given user group, and the second, i.e. subknowledge base, may be used only by a few in such a given user group. This enables different users to update their own knowledge base e.g. with special vocabularies or expressions, without such knowledge suitable for special purposes being used by other users.

In such a decentralised knowledge base, the updating of the first, i.e. the main knowledge base can be performed from the second, i.e. subknowledge bases. Data stored in the second knowledge bases are then transferred to the first knowledge base by predetermined criteria. One such criterion may be the incidence of specific data. The data exchange between the knowledge bases can also take place with one common main knowledge database administrator checking and approving each data to be transferred.

A number of embodiments of the solution in accordance with the invention has been described above. The principle of the invention can, of course, be varied within the scope of protection of the claims, for instance regarding details of the embodiment and fields of application.

Claims

1. A method for machine translation of information given as a character string in a first language into a character string in a second language, comprising
 - 5 - storage in the knowledge base of model segments in the form of character strings in said first language, and in logical connection with these, model segments (133, 134) in the form of character strings in the second language,
 - identifying a structural segment in the character string of said first language following a first rule (102),
 - 10 - comparing said identified structural segment with model segments (104) in the form of character strings in the first language stored according to a second rule,
 - striving to select one model segment (110) on the basis of said comparison,
 - reading a model, i.e. equivalent segment (121) in the form of a character string in the second language logically connected to the selected model segment, and
 - 15 - translating said structural segment into said translation segment in the form of a character string in the second language on the basis of said equivalent segment and a third rule (122),**characterised** in that the method comprises the identification of an intermediate word and/or a suffix and said first rule is essentially based on the identification of
20 said intermediate word and/or suffix.
2. A method as claimed in claim 1, **characterised** in that said information to be given as a character string in the second language is generated on the basis of translation segments and a fourth rule (124).
- 25 3. A method as claimed in claim 1 or 2, **characterised** in that, when no model segment to be selected following the second rule is found as a result of the comparison of the structural segments, the structural segment is displayed by means of a user interface (131) and the equivalent segment of the displayed structural segment is stored in the knowledge base by means of the user interface (132, 133).
30
4. A method as claimed in any of the preceding claims, **characterised** in that said structural segment comprises a punctuation mark.
- 35 5. A method as claimed in any of the preceding claims, **characterised** in that the type identifier of the model segment is stored in logical connection with the model segment.

6. A method as claimed in any of the preceding claims, **characterised** in that there are more than two model segments representing different languages logically connected to each other.

5 7. A method as claimed in any of the preceding claims, **characterised** in that one of said rules is updated on the basis of output data from the user interface.

8. A method as claimed in any of the preceding claims, **characterised** in that information is fed over the user interface to update the knowledge base with a view
10 to translate first information and said input data is used to update other data than those needed for the translation of said first information in said knowledge base.

9. A method as claimed in any of the preceding claims, **characterised** in that it comprises steps for

- 15 - reading the first information given as a character string in the first language,
- translating the first information given as a character string in said first language on the basis of data in the knowledge base into first information given as a character string in the second language to the extent allowed by the data available in the knowledge base,
20 - determining the additional data needed to complete the translation of the first information given as a character string in the first language into first information in the form of a character string in the second language,
- feeding said additional data in the knowledge base to update the knowledge base,
- completing the translation of the first information given as a character string in the
25 first language into first information given as a character string in the second language,
- storing said first information given in the second language,
- reading the second information given as a character string in the first language,
- translating the second information given as a character string in said first language
30 into second information given as a character string in the second language on the basis of said updated data in the knowledge base.

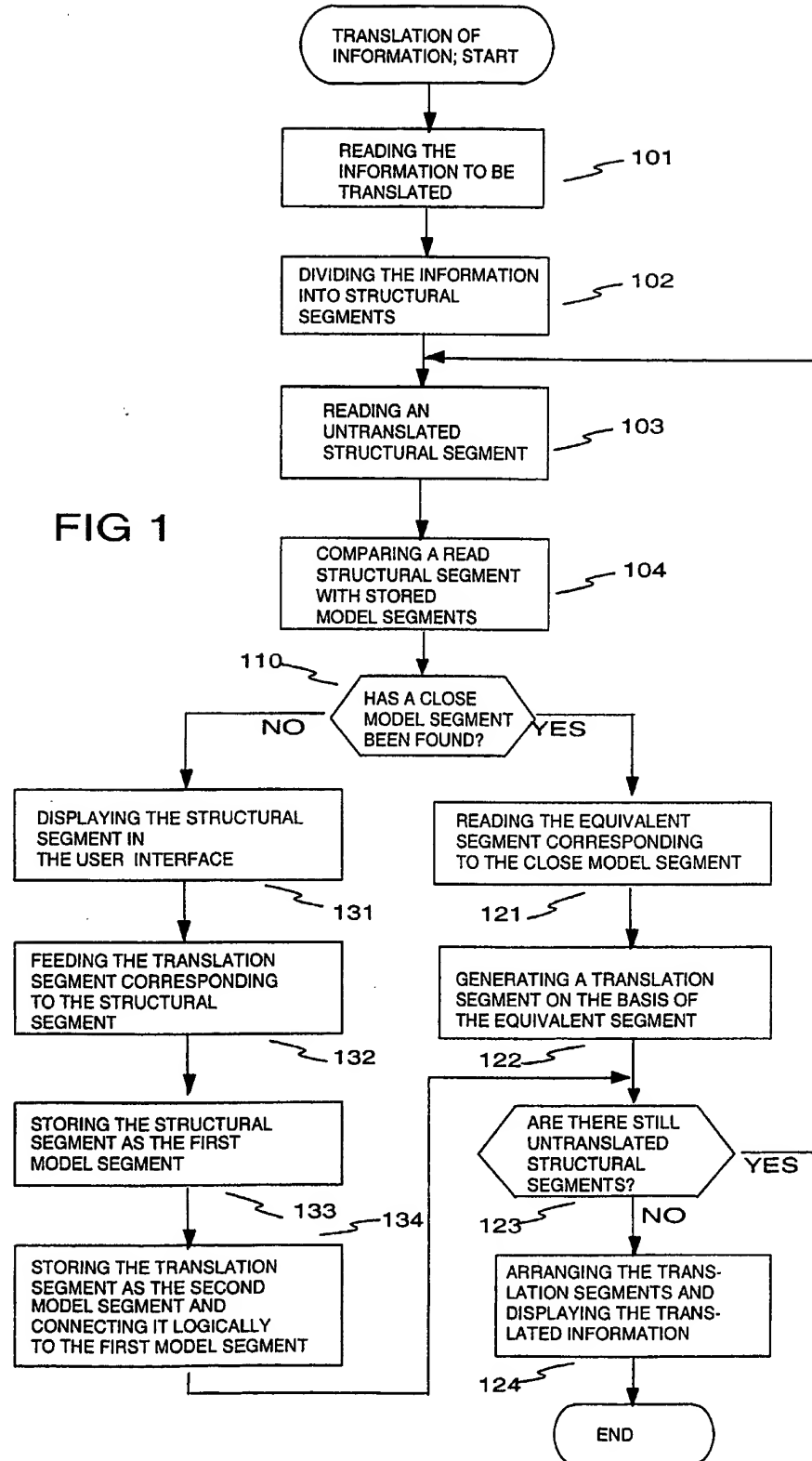
10. An arrangement for translating information given as a character string in a first language into a character string in a second language, comprising

- 35 - knowledge base means (20, 25) for storing model segments in the form of character strings in said first language, and in logical connection with these, equivalent segments in the form of character strings in the second language, and for storing a first, second and third rule,

- means (20, 24) for identifying structural segments in the information given as a character string in said first language following a first rule,
 - means (20, 25) for comparing said identified structural segment with the model segments stored in the form of character strings in the first language following a second rule,
 - means (20) for selecting one model segment on the basis of said comparison,
 - means (20, 25) for reading the model, i.e. equivalent segment in the form of a character string in the second language logically connected to the selected model segment in said knowledge base means and
 - means (20, 24) for translating said structural segment into said translation segment in the form of a character string in the second language on the basis of said equivalent segment and a third rule, said translation segment representing the information to be given in said second language,
characterised in that said means (20, 24) for identifying the structural segment in said information given as a character string in the first language comprise means for identifying an intermediate word and/or suffix, said first rule being essentially based on said identification of the intermediate word and/or suffix.
11. An arrangement as claimed in claim 10, **characterised** in that it further comprises means (20, 25) for generating information to be given as a character string in the second language on the basis of at least two translation segments and a fourth rule.
12. An arrangement as claimed in claim 10 or 11, **characterised** in that it comprises user interface means (22, 23) for connecting the user to said knowledge base means.
13. An arrangement as claimed in claim 12, **characterised** in that the user interface means are connected to said knowledge base means over a data transmission network.
14. An arrangement as claimed in any of claims 10-13, **characterised** in that said knowledge base means comprise a first knowledge base means (25) and a second knowledge base means so that specific users have access to said first knowledge base means and only some of said specific users have access to said second knowledge base means.
15. An arrangement as claimed in any of claims 10-14, **characterised** in that said knowledge base means comprise a first knowledge base means (25) and a second

knowledge base means, the arrangement comprising means for data input from the user interface means to said second knowledge base means and means for selective transfer of data stored in said second knowledge base to said first knowledge base means.

1/2



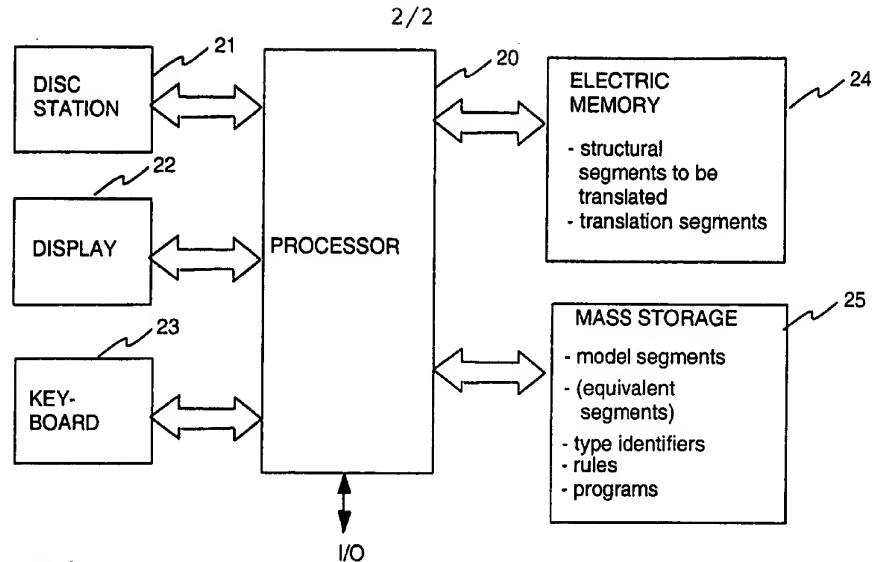


FIG 2

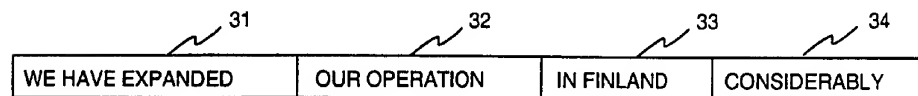


FIG 3

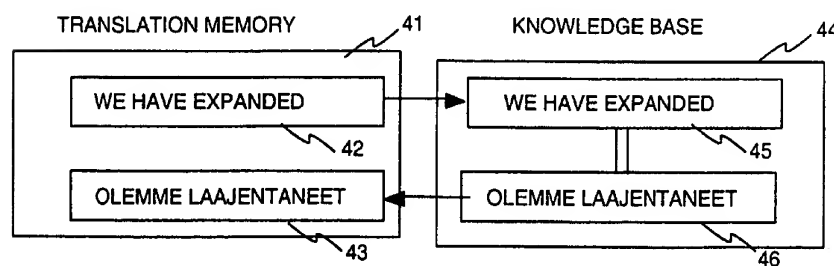


FIG 4

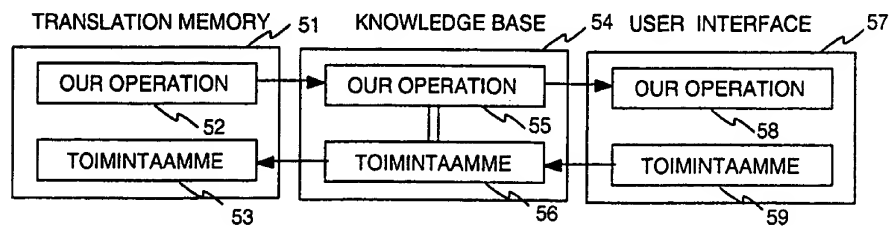


FIG 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 98/00441

A. CLASSIFICATION OF SUBJECT MATTER		
IPC6: G06F 17/28 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
IPC6: G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
SE,DK,FI,NO classes as above		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
WPI, JAPIO		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0805403 A2 (SONY CORPORATION), 5 November 1997 (05.11.97), see whole document --	1-15
A	EP 0262938 A1 (BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY), 6 April 1988 (06.04.88), see whole document -- -----	1-15
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "I" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
1 April 1999		08-04-1999
Name and mailing address of the ISA: Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. + 46 8 666 02 86		Authorized officer: Jan Silfverling Telephone No. + 46 8 782 25 00

INTERNATIONAL SEARCH REPORT

Information on patent family members

02/03/99

International application No.

PCT/FI 98/00441

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0805403 A2	05/11/97	JP 10021243 A	23/01/98
EP 0262938 A1	06/04/88	SE 0262938 T3	
		CA 1294056 A	07/01/92
		CN 1029170 B	28/06/95
		DE 3788488 D,T	11/05/94
		ES 2047494 T	01/03/94
		HK 136196 A	02/08/96
		IN 170322 A	14/03/92
		JP 1501977 T	06/07/89
		JP 2848458 B	20/01/99
		RU 2070734 C	20/12/96
		US 5384701 A	24/01/95
		US 5765131 A	09/06/98
		WO 8802516 A	07/04/88